REPORT RESUMES

ED 018 629 08 VT 003 315

THE DEVELOPMENT OF ACHIEVEMENT MEASURES FOR TRADE AND TECHNICAL EDUCATION. PROGRESS REPORT NUMBER TWO.
NORTH CAROLINA UNIV., RALEIGH, N.C. STATE UNIV.

REPORT NUMBER BR-5-1319 PUB DATE 66
GRANT OEG-2-6-000517-0585
EDRS PRICE MF-$0.25 HC-$1.80 43P.

DESCRIPTORS- *ACHIEVEMENT TESTS, *TEST CONSTRUCTION, *TRADE AND INDUSTRIAL EDUCATION, RESEARCH PROJECTS, TECHNICAL EDUCATION, CURRICULUM, PSYCHOMOTOR SKILLS, TACTUAL PERCEPTION, CURRICULUM EVALUATION, NORTH CAROLINA,

THE 11 INSTITUTIONS AGREEING TO PARTICIPATE IN THE STUDY TO DISCOVER MORE ADEQUATE WAYS OF MEASURING IMPORTANT TACTILE-KINESTHETIC MODALITIES WERE VISITED, AND THE OPERATING PROCEDURES AND INFORMATION ON WHAT WOULD BE EXPECTED OF EACH INSTITUTION WERE OUTLINED. APPROXIMATELY 20 INSTRUCTORS BEGAN WORK ON A DETAILED CURRICULUM ANALYSIS IN THEIR RESPECTIVE AREAS AND ON THE DEVELOPMENT OF ITEMS FROM WHICH A PRELIMINARY FORM OF AN EXAMINATION WILL BE MADE. A PRELIMINARY FORM OF A COMPREHENSIVE ELECTRONICS TECHNOLOGY EXAMINATION AND RESULTS FROM 41 STUDENTS WHO TOOK THE EXAMINATION 2 YEARS AGO WERE MADE AVAILABLE BY THE CURRICULUM LABORATORY, NORTH CAROLINA DEPARTMENT OF COMMUNITY COLLEGES. STATISTICAL ANALYSES WERE PERFORMED, AND IT IS FELT THAT THIS EXAMINATION MAY FORM A NUCLEUS FOR THE PAPER AND PENCIL TEST TO BE DEVELOPED. THE STAFF DECIDED TO ATTEMPT TO MEASURE TACTILE-KINESTHETIC SENSITIVITY BY AN EXPERIMENT UTILIZING THE PSYCHOPHYSICAL APPROACH. SPECIALIZED APPARATUS WAS DESIGNED AND CONTRACTED FOR CONSTRUCTION. THE APPENDIXES CONTAIN (1) AN AMENDMENT TO THE RESEARCH PROJECT, TO PROVIDE A BASIC RESEARCH EFFORT DIRECTED TOWARD OBTAINING ANSWERS TO SOME OF THE QUESTIONS EMANATING FROM THE ACHIEVEMENT TEST DEVELOPMENT, AND (2) WORKING PAPERS TITLED "TACTILE-KINESTHETIC SENSITIVITY EXPERIMENT" AND "THE ASSESSMENT OF TRADE AND TECHNICAL EDUCATION COURSES--THE USE OF INSTRUCTOR AND PEER RATINGS AS INTERMEDIATE CRITERIA OF CURRICULA EFFECTIVENESS." A BIBLIOGRAPHY IS INCLUDED. OTHER PROGRESS REPORTS ARE VT 003 205, VT 003 699, VT 005 199, AND VT 005 200. (HC)

BR-5-1319
PA 08

ED018629

OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

SUMMARY OF PROGRESS REPORT

During the reporting period all institutions requested to participate in the initial phase of the research project were visited. Operating procedures and what would be expected of each institution should it participate were also outlined. To date, eleven of the thirteen institutions have agreed to participate fully in the initial phase.

Approximately twenty instructors have already agreed to work on the project. These instructors have been placed under contract and have begun work on a detailed curriculum analysis in their respective areas. They are also beginning to develop items from which a preliminary form of an examination will be made.

A preliminary form of a comprehensive electronics technology examination which was developed by the Curriculum Laboratory, Department of Community Colleges, was made available to the project along with the results of forty-one students who took the examination two years ago. Statistical analyses were performed. It is felt at this time that this examination may form a nucleus for the paper and pencil test which we will develop.

A study of tactile-kinesthetic sensitivity was designed. Apparatus needed to conduct this study was designed and a contract for its construction was awarded.

A visit was made to the Instructional Materials Laboratory, Trade and Industrial Education, The Ohio State University, to discuss their achievement measurement activities and to insure a close coordination with their activities.

VT003315

PROGRESS REPORT
NUMBER TWO

THE DEVELOPMENT OF ACHIEVEMENT MEASURES FOR TRADE AND TECHNICAL EDUCATION

Office of Education Grant Number OEG 2-6-000517-0585
The Vocational Education Act of 1963, P.L. 88-210,
Section 4(c)

North Carolina State University at Raleigh
Raleigh, North Carolina

1 June-31 August 1966

The question of how best to obtain a curriculum analysis from each instructor was the object of many staff meetings. Two basic approaches were advanced. The first approach was that the material in each curriculum be broken down according to the psychological processes used by the student to learn it. The Taxonomy of Education Objectives, Cognitive Domain, by Benjamin Bloom, et al., represents the most elaborate attempt to use this basic approach. The project staff spent considerable time attempting to classify the behavioral objectives according to this system and several similar systems suggested by the project staff. It was found that considerable confusion existed among professional psychologists and educators in applying this system to the conversion of behavioral objectives into test items. It was concluded that such a broad conceptualization of behavior would not be a useful tool to provide subject matter experts with the expectation that they could write items appropriate for each of the taxonomic categories.

The second approach, which was the one finally adopted for use, is an inductive rather than deductive one. It calls for the instructors to do their curriculum analysis by writing actual test items which they feel will adequately measure the student's acquisition of all relevant material. In this way it was felt we would be starting off with a pool of items which give us the curriculum analysis we need. Further analysis into behavioral objectives will be accomplished primarily through statistical analyses.

During the reporting period all institutions requested to participate in the initial phase of the research project were visited. In each institution, the chief administrative officer was informed as to the objectives of the research program. Operating procedures and what would be expected of each institution should it participate were also outlined. To date, eleven of the thirteen institutions have agreed to participate fully in the initial phase. One institution has agreed to make students available to us, and one institution has been eliminated from the project because the institution lost instructors in each of the curricula in which we desired their participation.

To the extent possible, instructors in institutions involved were included in the explanation of the project and what would be expected of them should they agree to participate. Approximately half of the instructors which we desired to participate in the project were away at school or on vacation for the summer. These instructors will be contacted as soon as they return for fall quarter.

Approximately twenty instructors have already agreed to work on the project. These instructors have been placed under contract and have begun work on a detailed curriculum analysis in their respective areas. These instructors are also beginning to develop items from which a preliminary form of an examination will be made. Guidelines for curriculum analysis and item writing for a pool of test items were developed and sent to each of the instructors mentioned above.

A preliminary form of a comprehensive electronics technology examination which was developed by the Curriculum Laboratory, Department of Community Colleges, was made available to us along with the results of forty-one students who took the examination two years ago. To date, an item analysis, reliability of subtests and correlation between subtests has been performed. It is felt at this time that this examination may form a nucleus for the paper and pencil test which we will develop.

A meeting was held in Raleigh on June 28, 1966, with Dr. Rupert N. Evans, a consultant on the project, and Dr. Otto P. Legg, Project Monitor.

The project aims, along with general and specific problems were discussed, and in some cases revised and broadened. At this meeting the possibility of submitting an amendment to the original contract was discussed. The additional work to be conducted under the amendment would deal with some of the basic research problems involved in the measurement of non-paper and pencil aspects of achievement. A copy of the amendment is enclosed as Appendix A. The research would be done currently and integrated with the original aims of the project.

Possible new locations for the project staff have been investigated and several alternatives presented to the administration. The new site would contain three soundproof experimental rooms which are essential to the basic research interests of the project. The several alternatives are now being considered by the administration.

Trips were made to Washington and New York to consult with people in the military about the area of psychomotor research. The consensus of opinion was that very little of this type of work was going on in the area of achievement measurement at present. However, past work in the military has dealt largely with testing for aptitude and not achievement. The majority of people, while showing much interest in our project, were unable to be of much assistance since little or no work has been done in achievement testing.

In accordance with our belief that the tactile-kinesthetic sense was of utmost importance to skill development, staff discussions were held to determine the best way of measuring this. Finally an experiment utilizing the psychophysical approach was decided upon. It is to be noted that while a psychophysical approach is being used to conduct the experiment, the classical psychophysical method of viewing individual differences as error variance is not being subscribed to. On the contrary, these individual difference scores are the area of most interest to us. A summary of the experimental design and proposed data analysis is included in Appendix B. Much discussion and preliminary design were necessary before the finalized version of the apparatus was completed. Unfortunately some time was lost in the process of designing and building the specialized apparatus to cope with a problem. A delivery date of October 1, for the completed apparatus appears realistic.

Mr. William L. Ballenger, a product design consultant, was hired on a consulting basis to design the apparatus which consists of three separate units. Each unit is functionally identical, differing only in the plane in which it will move, i.e., vertical or horizontal.

Components of each unit are a 28 inch metal rod machined to accept a friction-free linear ball bearing. The bearing is fitted with an adapter so that a lever can be attached. Through a system of pulleys, it is possible to attach counterweights to vary the resistance of the lever to movement. By manipulating the counterweights attached to the apparatus, it will be possible to study individual tactile-kinesthetic sensitivity as discussed in Appendix B.

Dr. Norman E. Stander spent some time visiting the project on a consulting basis. Dr. Stander worked with the project staff on problems of experimental design, psychophysical methods, and secondary criteria that might be used. A copy of the report prepared by Dr. Stander on the last topic is included as Appendix C.

A meeting of the Steering Committee was held on June 15, 1966, to provide the committee members with a progress report and to solicit their suggestions.

A visit was made to the Instructional Materials Laboratory, Trade and Industrial Education, The Ohio State University, to discuss their achievement measurement activities. While there are common areas of interest, their program is primarily designed for high school level programs. Contact will be maintained with this organization in an effort to exchange technical data.

Mr. Chris C. Y. Hsu joined the project staff on September 1, 1966 as a research assistant.

APPENDIX A

PROPOSAL FOR RESEARCH PROJECT

Submitted to the United States Commissioner of Education
under the Provisions of Section 4 (c)
of the Vocational Education Act of 1963

Project Title:	Amendment to The Development of Achievement Measures for Trade and Technical Education Grant No. OEG-2-6-000517-0585
Submitted by:	North Carolina State University School of Education Department of Industrial Education
Address:	Tompkins Hall, Hillsboro Street, Raleigh
Telephone Number:	Area Code 919 755-2241
Initiated by:	Thomas S. Baldwin, Research Associate Professor Department of Industrial Education Tompkins Hall, Hillsboro Street, Raleigh Telephone: Area Code 919 755-2241
	H. F. Robinson Administrative Dean for Research North Carolina State University at Raleigh
Submitted by:	A. H. Shepard, Jr. Assistant Vice President and Treasurer University of North Carolina at Chapel Hill
Federal Funds Requested:	$116,811
Duration:	One year, beginning with effective date of contract
Date Transmitted:	August, 1966

AMENDMENT TO RESEARCH GRANT OEG-2-6-000517-0585 (Project No. ERD-517)

Section I - Introduction

During the first several months of this research project, it has become apparent that there are many questions which need to be answered through a basic research effort. The research contract itself does not propose to engage in basic research, but rather is directed toward the applied problem of developing achievement measures in several trade and technical programs. This amendment is intended to provide a basic research effort, closely coordinated with the applied problems with which the research grant deals, which will be directed toward obtaining answers to some of the questions which are brought to light in the course of developing the achievement tests.

Section II - Rationale

This section is an attempt to elaborate on some of the typical problems which have come to light as a result of the initial efforts on this project. The problems mentioned here should be considered illustrative of the type of work to be conducted under the proposed amendment since other problems undoubtedly will require such research effort.

Psychologists on this project have devoted considerable attention to the question of how the non-cognitive behaviors (for example, psychomotor performance, auditory discriminations, etc.), might be abstracted from the job and measured objectively in an institutional setting. The first and most obvious means of measuring such behaviors is to have the individual actually perform the job, or a representative sampling of tasks which comprise the job. This approach is essentially what is done in many training and evaluation situations and is exemplified by the NASA approach to astronaut training and proficiency measurement. With this approach one recreates the "real world" in which the individual will later be expected to perform and has him perform appropriate work tasks. This may vary in complexity from an auto mechanic actually doing a tune-up on an automobile to an astronaut flying a simulator around a man-made moon.

While the approach is a very "high fidelity" one and has a great deal of face validity, it is our belief that this is not necessarily the only way in which one might measure the appropriate behaviors. It is our feeling that behaviors could be abstracted from the real world situation and measured rather than having them embedded in an actual job context. Such an approach is essential to implementing the results of this project since whatever achievement measures are derived will have to be of sufficient ease in administration, scoring, and in general logistics that they can be used in an institutional setting by personnel essentially unskilled in sophisticated techniques of psychometrics. The approach of abstracting the job behaviors that our curriculum analysis shows to be important has the other very distinct advantage over the work sample approach of providing a more general solution to measuring job proficiency which should be applicable to fields other than the ones in which we are directly involved.

Considerable attention has been given to the question of how we will measure psychomotor behavior. (It should be mentioned at this point that other aspects of behavior are also important, such as auditory sensitivity, visual sensitivity, etc., but psychomotor behavior is being used here for illustrative purposes). The question of how to measure the cognitive aspects of a job is one which, from a methodological point of view, has been dealt with effectively by psychologists for some time. The "new ground to be plowed" in this project, however, has to do with the non-cognitive aspects of the job. For example, our original conception of the psychomotor component was that we were dealing with the output or motor aspect of behavior. After reviewing the literature and discussing the question among ourselves, we have tentatively concluded that it is not the motor or output side of behavior that undergoes a major change as a function of training, but rather it is the input side that is modified. This is true at least within the context of those jobs with which we are working. As we presently conceive the problem, we don't feel that individuals undergoing training in programs with which we are dealing acquire large numbers of new responses, although they may put together certain responses that are already in their repertoire in new combinations.

It is certainly true that in some training responses may be modified. For example, the response made by a boxer would not be the same before he began training and after he became proficient at boxing, since substantial changes in the musculature which determine his response would have occurred. We do not feel that similar changes in the "output mechanisms" of the human occur within the training courses that we are discussing or in fact within most training. For example, the response necessary to depress the brake pedal in an automobile is present even in people who have never driven an automobile before. Learning to drive an automobile does not then give that person a new response of pressing the brake pedal with the right foot. If this argument is true, what then occurs during training which makes a man a proficient driver or a proficient machinist and distinguishes him from the untrained individual? At this point in our thinking, we feel that the major behavioral changes which occur are on the input side rather than the output side. In other words, the person learns to make the same response to different stimuli.

To cite another example, take the case of a mechanic using feeler gauges to adjust the valves or points on an automobile; this would traditionally be described as a psychomotor performance. However, the physical manipulation of the feeler gauges is not a new response. Rather, what he is learning in training is to make very fine sensory discriminations, such as being able to feel very small degrees of friction exerted on the gauge. While other things may be involved as well this would probably represent the major portion of his learning. If this analysis is correct, it would follow that our attention should be devoted to the changes that occur in an individual's ability to sense very small differences in input.

Section III - Proposed Work

The work to be conducted under this amendment will be to study individual differences on such variables as proprioceptive feedback, tactile sensitivity, auditory sensitivity, etc., as they are modified by training and contribute toward achievement in the several trade and technical curricula with which the grant deals. It is our opinion that these variables can be isolated and measured and that they will adequately represent the non-cognitive tasks with which the worker is confronted, so that it will not be necessary to have him perform the actual job. Our intention then is to undertake research studies of the dimensions of proprioceptive feedback, tactile sensitivity, auditory sensitivity, etc., and how these change as a function of training. In reviewing the literature it was interesting to note that the importance of this fact was mentioned by several authors and that some rather isolated attempts had been made to exploit the factor of individual differences on such things as kinesthetic sensitivity.

Fitts[1] suggests, for example, that what happens in learning a motor skill is that the individual learns to respond to interoceptors rather than exteroceptors. In the unskilled individual, cues such as vision are the input to which the individual responds. In the skilled worker, however, it is the proprioceptive cues to which the individual responds. In a study reported by Fleishman[2] he empirically demonstrated the validity of this hypothesis. The study by Fleishman is an interesting example of what we feel is the most profitable line of research. He was able to demonstrate rather marked findings using a measure of kinesthetic sensitivity (sensitivity in lifting weights) which we feel is inadequate for describing the entire picture of the kinesthetic sense modality. A more adequate description of kinesthesis would seem to have considerably more promise for predicting behavior on psychomotor tasks. The measure of kinesthetic sensitivity used by Fleishman might be considered analogous to describing a person's total visual capabilities in terms of his visual acuity. In other words, while Fleishman's study adequately supported his hypothesis, it is felt that the measure of kinesthetic sensitivity was less than adequate for describing in its entirety the kinesthetic sense modality. If a more adequate description of this modality were available, it is reasonable to expect that the prediction of motor performance would be much more accurate.

---

[1] Fitts, Paul M., in Handbook of Experimental Psychology, John Wiley & Sons, New York, 1951, p. 1323-4.

Our present thinking is that this is a most fruitful line of research to pursue. While we intuitively feel that the kinesthetic sense modality is of considerable importance in some of the areas in which we will be working, we are not sure to what extent other sense modalities will come into play. Certainly on a logical basis one would think that audition is important in some situations. For example, an auto mechanic would probably use auditory cues in performing diagnostic work on an engine. It might well be that the skin senses are of considerable importance to a machinist. The extent to which the other modalities (other than the kinesthetic) come into play in these various jobs is being determined by our curriculum analyses and basic studies of how these modalities influence achievement in trade and technical training would be undertaken under the proposed amendment.

---

[2] Fleishman, E. A. and Rich, S. Role of Kinesthetic and Spatial-Visual Abilities in Perceptual-Motor Learning. _J_. _Exp_. _Psychol_., 1963, 66, 6-11.

An initial experiment is proposed to derive a more adequate measure of kinesthetic sensitivity. If it were possible to determine the orthogonal dimensions of kinesthetic sensitivity in arm and hand movement, it would then be possible to study what happens to this sensitivity as a function of training. This would permit us to study the hypothesis that it is primarily the input function which becomes modified during training. The implications of an understanding of how this occurs would seem to be considerable, not only from the standpoint of developing achievement tests, but also for other considerations such as the design of training programs. Essentially we are interested in discovering the dimensionality of kinesthetic sensitivity, at least insofar as the arms and hands are concerned. And it would be individual differences, the stability or reliability of these differences, the orthogonality of the dimensions with which we would be most concerned. The methodology that we would propose to use at this point is one of studying difference limens in a number of movements to determine sensitivity of individuals and to determine the correlation between difference limens for varying types of movement. This is an approach that, to our knowledge, has never been taken by the psychophysicists, who are of the classical experimental school, and who treat individual differences as error variance. It would seem reasonable to hypothesize for example, that if one had measures of difference limens for movements in a number of directions, a general factor of kinesthetic sensitivity would emerge. This might be accounted for by the general state of development of the kinesthetic receptors and might be thought of as analogous to the general state of muscular development among people. People with large leg muscles tend to have large arm muscles, etc. It would seem reasonable also to hypothesize sub-general factors of kinesthetic sensitivity which might reflect the involvement that a particular set of proprioceptors had in a number of different movements. There might also be a number of highly specific kinesthetic factors. In any event, it would seem that Fleishman's original attempt at using kinesthetic sensitivity and studying its interaction with proficiency over training trials could be more adequately done if a broader understanding of the kinesthetic sense modality were available. The experimental design for an initial experiment to be conducted under this amendment is enclosed as Appendix A.

Section IV - Budget

	Grant Funds	Applicant or Other Funds
Direct Cost		
Personnel		
1. Psychologist (EdD, full time, 12 months)	$ 12,500	$
2. Shop Technician (Half time, 12 months)	3,000	
3. Graduate Assistants (2 @ $2,000)	4,000	3,000
4. Matching Contributions (11% of $15,500)	1,705	
Supplies and Materials		
5. Telephone ($20.00 per month)	240	
6. Postage ($5.00 per month)	60	
7. Psychological Test Apparatus	8,000	
Other		
8. Staff Travel	500	
Total Direct Cost	$ 30,005	$ 3,000
Overhead @ 34.9%	6,806	1,047
Total Costs	$ 36,811	$ 4,047
Requirements beyond first year		
Amount required for second year	$ 40,000	$ 4,047
Amount required for third year	$ 40,000	$ 4,047
Total	$ 116,811	$ 12,141

APPENDIX A

TACTILE-KINESTHETIC SENSITIVITY EXPERIMENT

It is believed that the tactile-kinesthetic sense plays a major role in psychomotor skill development. This experiment is the first of a series which is being designed to provide some insights into the dimensionality and reliability of the tactile-kinesthetic sense modality. This will be done by determining the weight sensitivity difference limens (D.L.) for three different positional arm, wrist and finger movements, in eight distinct directions. The experiment will yield a total of 24 D.L. per subject. The different arm positions movement variables are shown in Table I.

TABLE I.

	Arm Position	Movement Variable		
(1)	Full extension (maximize use of shoulder)	1.	Vertical	(Center-up)
		2.	Vertical	(Center-down)
		3.	Horizontal	(Center-right)
		4.	Horizontal	(Center-left)
		5.	Depth	(Center-aft)
		6.	Depth	(Center-fore)
		7.	Roll	(Clockwise)
		8.	Roll	(Counterclockwise)
(2)	Half extension (maximize use of forearm and wrist)	1.	Vertical	(Center-up)
		2.	Vertical	(Center-down)
		3.	Horizontal	(Center-right)
		4.	Horizontal	(Center-left)
		5.	Depth	(Center-aft)
		6.	Depth	(Center-fore)
		7.	Roll	(Clockwise)
		8.	Roll	(Counterclockwise)
(3)	Zero extension (maximize use of fingers and wrist)	1.	Vertical	(Center-up)
		2.	Vertical	(Center-down)
		3.	Horizontal	(Center-right)
		4.	Horizontal	(Center-left)
		5.	Depth	(Center-aft)
		6.	Depth	(Center-fore)
		7.	Roll	(Clockwise)
		8.	Roll	(Counterclockwise)

The use of different arm positions and movement variables will allow us to determine to what extent tactile-kinesthetic sensitivity in one movement is correlated with that sensitivity in other movements.

Experimental Design

Initially the subjects used in this experiment will be students enrolled in the introductory psychology classes at the university. We do, however, intend to use students in the community college after our pilot studies are complete.

The apparatus for this study will consist of three 27" x 6" metal plates with a 3/8" groove down the center of each. A lever which can be moved across the groove will protrude about 2" from the center of each groove. Each metal plate will be mounted in a different position (vertical, horizontal, and depth) to facilitate the necessary experimental movements. On each trial the subject will be asked to move the lever in the desired direction along the groove. On the opposite side of each plate, out of the subjects view, different size weights will be connected to the lever by the experimenter according to the experimental procedure. In order to facilitate the different arm positions, the subject arm will be placed in a restraining device which will be attached to his chair.

Data Analysis

The data yielded by this study will consist of 24 measures of sensitivity for each of the subjects in the experiment. The intercorrelations for these 24 measures will be computed, yielding a 24 by 24 matrix of intercorrelations. This matrix will reflect the extent to which tactile-kinesthetic sensitivity in one movement is correlated with that sensitivity in other movements.

In order to determine the underlying dimensionality of tactile-kinesthetic sensitivity, the matrix will be subjected to a centroid factor analysis and factors extracted until the matrix is reduced to error.

While it is possible to hypothesize the existence of several factors on rational grounds, the entire factor structure cannot be anticipated since no previous experimental work exists in this area. The first factor hypothesized is a general factor on which all 24 variables should load. This factor would represent the general state of development of the tactile-kinesthetic sense modality. Group factors, reflecting the involvement of the same muscle group in several different movements, are hypothesized. It is also hypothesized that specific factors will appear where a given muscle group is involved in only one movement. Graphic and mathematical rotations of the factor loadings will be made with these factors in mind.

Test re-test data will be obtained for each of the 24 dimensions of kinesthetic-tactile sensitivity. The time interval will be approximately 24 hours. Pearson product-moment correlation coefficients will be computed for each movement to yield a measure of reliability. Specificity will be estimated by comparing the communality to the reliability.

APPENDIX B

WORKING PAPER NO. 2 - TACTILE-KINESTHETIC SENSITIVITY EXPERIMENT

Objectives

The present experiment has two major objectives.

(1) To determine the dimensionality of the tactile-kinesthetic sense, through the use of different positional arm, wrist, and finger movements.

(2) To determine the reliability of the tactile-kinesthetic sense. The Ss will be retested 24 hours after the first test. The test re-test reliability method will be used in this situation.

Experiment

Subjects

Subjects will be male students in the Psychology 200 classes in North Carolina State University. The number of Ss will be more than 30.

Dependent Variables

The classical approach of psychophysics, namely the method of limits, will be used in this experiment. The dependent variable will be the difference limen (D.L.) of the tactile-kinesthetic sense totaling in all 24 D.L. per S.

Independent Variables

The independent variables will be as follows:

	Arm Extension	Movement Variables
(1)	Full extension (stretch whole arm and hand)	1. Vertical (Center-Up) 2. Vertical (Center-Down) 3. Horizontal (Center-Right) 4. Horizontal (Center-Left) 5. Depth (Center-Aft) 6. Depth (Center-Fore) 7. Roll (Clockwise) 8. Roll (Counter-Clockwise)
(2)	Half extension (attach elbow to the trunk)	Ditto
(3)	Zero extension (attach elbow to the trunk, pull back the forearm as much as possible, move wrist and finger)	Ditto

The reason for these three positions, i.e. full extension, half extension, and zero extension is that we would like to separate the unit function of hand movements. Movements by full extension of upper arm, forearm, and hand, would be considered as maximum use of shoulder. Attaching the elbow to the trunk and moving both forearm and hand is called half extension, which would maximize the use of the forearm and wrist. Zero extension refers to the elbow attached to the trunk and the forearm pulled back as much as possible. This would probably limit the movements to finger and wrist motions. The present classifications are arbitrarily defined.

Experimental Design

Apparatus

The apparatus consists of a board (size 2 x 2 ft^2), on which are two grooves in the shape of a cross. Subjects will be asked to move a lever inserted in the grooves. On the opposite side of the board, weights will be connected to the lever. The experimenter can adjust the weights according to the experimental procedures. This same apparatus can be designed vertically or horizontally. For the measurements of depth movements, we can use the horizontal apparatus, while for the measurements of vertical and horizontal movements, we will use the vertical apparatus. In order to measure roll movements, the lever can be exchanged for a knob, which may be rolled clockwise or counterclockwise. See Figure I.

Procedures

All trials will be randomized and counterbalanced for arm extensions and movement variables, as well as for comparison stimuli presentation. The number of trials, distance of movements, increments of comparison stimuli, distance of apparatus from S and length of time for each S, etc., will largely depend on the outcome of a pilot study with the proposed apparatus.

Data Analysis

The data yielded by this study will consist of 24 measures of sensitivity for each of the subjects in the experiment. The intercorrelations for these 24 measures will be computed, yielding a 24 by 24 matrix of intercorrelations. This matrix will reflect to which tactile-kinesthetic sensitivity in one movement is correlated with that sensitivity in other movements.

In order to determine the underlying dimensionality of tactile-kinesthetic sensitivity, the matrix will be subjected to a centroid factor analysis and factors extracted until the matrix is reduced to error.

While it is possible to hypothesize the existence of several factors on rational grounds, the entire factor structure cannot be anticipated since no previous experimental work exists in this area. The first factor hypothesized is a general factor on which all 24 variables should load. Furthermore, this factor would represent the general state of development of the tactile-kinesthetic sense modality. Group factors, reflecting the involvement of the same muscle group in several different movements, are hypothesized.

It is also hypothesized that specific factors will appear where a given muscle group is involved in only one movement. Graphic and mathematical rotations of the factor loadings will be made with these factors in mind.

Test re-test data will be obtained for each of the 24 dimensions of kinesthetic-tactile sensitivity. The time interval will be approximately 24 hours, Pearson product-moment correlation coefficients will be computed for each movement to yield a measure of reliability. Specificity will be estimated by comparing the communality to the reliability.

APPENDIX C

WORKING PAPER NO. 3

The Development of Achievement Measures for Trade and Technical Education

The Assessment of Trade and Technical Education Courses: The Use of Instructor and Peer Ratings as Intermediate Criteria of Curricula Effectiveness

Norman E. Stander, PhD
Consultant

2 August 1966

WORKING PAPER NO. 3

The Assessment of Trade and Technical Education Courses: The Use of Instructor and Peer Ratings as Intermediate Criteria of Curricula Effectiveness

I. Overview

The ultimate success of any training program may be determined only by observing whether its graduates successfully perform the job duties for which they have been trained. Any criterion other than that which is based on the observed presence or absence of certain operationally defined, critical (Flanagan, 1954), job behaviors merely approximates the degree to which training is likely to be effective. In addition, because an individual is usually assigned to a specific job after he has been trained it is often not possible to assess the ultimate worth of his training for several months or longer. Aside from certain theoretical considerations (Thorndike, 1949) regarding the designation and definition of ultimate criteria such a situation is often neither experimentally desirable nor administratively feasible.

It becomes necessary, therefore, to approximate the hypothetically "true" order of success of an activity by developing intermediate criteria which, through rational or empirical analysis, have demonstrated a high degree of relevancy to the ultimate criteria. Methodologically, this can be best accomplished in four alternative although not mutually exclusive ways. First, identify a representative sample of behaviors which are unbiased reflections of the total universe of behaviors necessary for success in a particular activity and apply unitary weights to the resultant behaviorgrams (Flanagan, 1954; Wherry, 1950). Second, assuming that the intermediate criteria selected - whether traits or observable behaviors -

are not highly intercorrelated, enlist judges familiar with the performance characteristics to be assessed to rate the extent to which each characteristic contributes toward successful performance in the activity. Appropriate differential weights are then assigned to each performance sub-element (Toops, 1944).

A third approach, useful regardless of the degree of relationship which exists among intermediate subcriteria also relies on rating but neither requires trained judges nor judges familiar with a specific activity. Moreover, not more than 10-15 judges are necessary to obtain satisfactory results. In this approach a large universe of descriptive statements are scaled by the psychophysical method of equal appearing intervals and are assigned weights in such a manner that the continuum representing the psychological attribute under consideration is divided into equal interval units of measurement (Thurstone, 1928; Uhrbrock, 1950; Uhrbrock, 1961). In this regard, it is important to note that other scaling techniques may also be employed (Torgerson, 1958).

Finally, an appropriate intermediate criterion is any measure which has demonstrated repeatedly high correlation with later measures of ultimate success for the behavior or group of behaviors under investigation. An example of such an intermediate criterion is peer ratings which have shown a high degree of relevancy with respect to a wide variety of complex behaviors or activities as well as in a diversity of situations (Hollander, 1965; Fuchs et al, 1953; Weitz, 1958; Klieger et al, 1962; Prien et al, 1965).

Aside from being relevant (Nagle, 1953), it is also essential that the intermediate criteria used to approximate the ultimate criterion are reliable and unbiased (Brogden et al, 1950). Criterion bias will be introduced if precautions are not taken to prevent one or more of the following four conditions from occurring: (1) omission of elements from the intermediate criterion which are essential for eventual successful performance, (2) inclusion of elements in the intermediate criteria which do not exist in the ultimate criteria, (3) improper weighting of subcriteria comprising the intermediate criteria, or (4) using criterion scale units which are unequal. The presence of any of these conditions will distort predictor-criterion validities as well as partial regression weights, increase the error of measurement of the criterion, or may adversely effect criterion reliability.

II. Development of Criterion Measures for Trade and Technical Education Courses

There is little question that the most precise estimate of the ultimate criterion of a vocational training program derives from an analytic and objective determination of those behaviors which are necessary for successful job performance. This method not only minimizes intermediate criterion deficiency and contamination but also minimizes criterion distortion. (This method does not require the assignation of differential weights). Moreover, it precludes the introduction of error variance which may accrue as a result of unequal criterion scale units since scaling per se is not a requisite of this technique. (The only demand on the rater is whether the behavior has been observed or not, i.e., p is either equal or not equal to 1.0).

It becomes possible then, within certain limits, to accurately predict whether a trainee will successfully perform the job for which he has been trained by developing an instrument (achievement test) which contain a series of behaviorgrams that are known to be incorporated in the ultimate criterion. Predictive validity will be maximized to the extent that the course curricula shape and stimulate the trainee to emit the desired behaviors, and to the extent that the critical terminal behaviors have been correctly identified.

However, in the absence of any definitive knowledge of these terminal behaviors, it is not possible to develop the appropriate achievement tests. As a result, the ultimate criterion may be estimated only by developing intermediate criteria other than those which have behavioral dimensions.

It should be noted parenthetically, that intermediate criteria are nothing more than predictors of the ultimate criteria. The accuracy of the prediction will, of course, be a function of the degree to which the intermediate criteria are relevant to the ultimate criterion. Unfortunately, there is no technique by which to measure the extent to which the intermediate criteria are relevant.

Notwithstanding this limitation, it has been shown (Goertzel, 1941; Ghiselli et al, 1957; Barrett, 1966) that in many instances ratings, assuming they are properly conceived and executed, are satisfactory approximations of ultimate job performance. A more cogent argument for their use, however, is that until such time that behaviorally oriented achievement tests are developed, no other suitable predictors of a trainee's potential job success are available. It is for this reason that instructor and peer ratings are recommended as supplementary criteria for determining the eventual job success of trade and technical course trainees.

III. Instructor Ratings

A. General

Instructor ratings have long been used as intermediate criteria of student success both in vocational and secondary education (Bell, 1949; Broadhurst, 1949; Fleming, 1938; Kirkpatrick, 1959; Novak et al, 1951; Hayes, 1963). In those instances that instructor ratings have shown poor prediction validity it has been subsequently found that one or more of the following conditions were present: (1) the raters were poorly trained, (2) the behaviors they were required to rate were complex and poorly defined, e.g., personality traits, (3) the raters lacked specific instructions, (4) not enough time was set aside for the ratings, (5) the raters were rated under different conditions and against different criteria, (6) constant errors of rating were not obviated, or (7) the rating instrument was unreliable.

As implied above much of the criterion error variance generated by rating techniques (and consequent attenuation of predictive validities) may be minimized by using a rating technique which has proven reliability, well-defined or accurately scaled items, precludes or reduces constant errors, does not have as a prerequisite that the raters be highly trained, and requires little rater instruction and time to complete.

While no single rating system embodies all these criteria, the check-list method most nearly meets these requirements. With respect to predicting vocational success, the check-list format more than any other rating technique also has the added advantage of being a reliable measure of those nonmotor performance factors that, as pointed out by Patterson[1] (1956) appear to be important determiners of vocational success. Moreover, because its items are often scaled by the method of equal appearing intervals, the check-list technique minimizes criterion scale unit bias often found in ranking methods and constant errors invariably found in adjectival rating scales. A check-list is also considerably less expensive and time consuming to develop than a forced-choice rating instrument.

---

[1] ...This means that in attempting to predict success in training for, or performance in, a skilled trade, measures of manual ability or dexterity are not as important as ability to acquire technical knowledge, ability to know when and where to apply it, with appropriate skill, and ability to understand and plan a process or job.

Finally, it has been shown (Richardson, 1933; Knauft, 1948) that carefully constructed check-list rating scales have demonstrated reliabilities as high as .80.

B. Development of an Instructor Check-List Rating Scale for Trade and Technical Trainees

One reason why the check-list format has not had a wider use as a rating technique is that it necessitates the scaling of a large number of descriptive statements. This scaling process, while less arduous than the development and matching of discrimination and preference indices associated with the forced-choice technique, nevertheless represents a great investment of time and effort. However, since Uhrbrock's (1961) development of 2000 scaled items, this objection has been largely obviated.

Uhrbrock's statements refer mostly to individuals in industrial settings. However, because of their generality many of them are appropriate for assessing trade or technical course trainees. Moreover, as noted by Ghiselli (1957) check-lists are most useful for generating measures of general proficiency rather than measures of specific strengths and weaknesses. Therefore, although it is desirable it is not essential that check-list items be developed to assess a specific activity or ratees in particular situations.

Hence, the decision to select a statement for inclusion in the to-be-developed Trade and Technical Trainee Check-List (TTTCL) must depend on three other factors. They are: appropriateness, the statement's mean value, and the statement's variability as measured by its standard deviation. Other factors to consider are the number of items to be included in the check-list, the order in which the items are to be presented to the rater, and the instructions the rater should receive.

A preliminary examination of the 2000 Uhrbrock items has indicated that 149 are appropriate for inclusion in the TTTCL. These items are listed in Appendix I. In this context, an "appropriate" item is one which is obviously applicable to technical and trade trainees, is, with few exceptions, representative of a class of items with the same mean values and, again with few exceptions, has a standard deviation which is equal to or less than the first quartile standard deviation of the total 2000 item range of standard deviations.

In this regard, the median standard deviation reported by Uhrbrock is 10.53. Q_1 equals 8.82 and Q_3 equals 11.30. It is interesting to note that the 2000 standard deviations are somewhat negatively skewed ($Q_2 - Q_1 = 1.71$, $Q_3 - Q_2 = 0.77$) suggesting that the judges were able to agree more consistently about which statements constituted positive attributes than negative attributes.

The TTTCL is to be developed from the 149 items listed in Appendix I. It is important to note that regardless of which items are finally selected that wording must be exactly as shown by Uhrbrock since even a minor change in an item's semantic structure changes its average scale value and standard deviation.

Regardless of the speed with which check-list ratings may be completed, it must be kept in mind that the average instructor will have 15-20 students to rate. For this reason the length of the TTTCL should not exceed 50 items nor be less than 40 items. In this regard, it can be shown that if the instrument has a reliability of .70 doubling its length would increase it to .82. Moreoever, if the original reliability was .90 doubling its length would increase the reliability to only .95.

Insofar as which specific items to select for the TTCL is concerned, it is essential that the sample of items selected (whether 40 or 50) represent the whole range of mean scale values reported by Uhrbrock. This range is from 10.00 to 109.38.

Traditionally, assuming 50 items and 5 scale intervals, this would require that the items be normally distributed as shown in Table I.

TABLE I.

Normal Distribution of 50 Items from Uhrbrock's 2000 Scaled Items

Decile	Range of Mean Scale Values	Number of Items to be Selected
1	10.00 - 24.99	3
2 - 3	25.00 - 37.49	13
4 - 6	37.50 - 83.74	18
7 - 8	83.75 - 95.62	13
9	95.63 -109.38	3

It is recommended, however, that to increase the scale's potential variance and, hence, the sensitivity of the TTTCL to distinguish among less competent and more competent trainees, that the items be distributed rectangularly as shown in Table II.

TABLE II.

Rectangular Distribution of 50 Items from Uhrbrock's 2000 Scaled Items

Decile	Range of Mean Scale Values	Number of Items to be Selected
1	10.00 - 24.99	10
2 - 3	25.00 - 37.49	10
4 - 6	37.50 - 83.74	10
7 - 8	83.75 - 95.62	10
9	95.63 -109.38	10

Although the check-list items are to be distributed rectangularly, it is important to note that the item scale values should, insofar as possible, be distributed normally within each of the five intervals defined by the decile ranges indicated in Tables I and II. In those instances in which two or more items have the same scale values and are equally acceptable, the item with the smallest standard deviation should be selected.

When the final items are chosen they should not be presented to the rater in an ordered array, for example, from highest to lowest or vice versa. Each item should be given a randomly assigned position. This method of presentation will reduce the raters' response set, minimize habituation errors, and increase the possibility that the rater will read all the items.

The instructions to the raters should be brief, simple, to the point, and consistent from rater to rater and in different situations. This is important since at least one study (Fisk et al, 1960) has reported that ratings tend to vary according to rating instructions and trait descriptions. It is also recommended that no restriction be put upon the instructors regarding the number of items on which the trainees should be rated (Ghiselli et al, 1957). A trainee's evaluation score will be the average of the scale values of the items checked by the instructor.

In the event that a majority of instructors check less than, say, 7-10 items on the TTTCL it will be necessary to change the "no restriction" instruction. In this regard, it is recommended that a small pilot study involving perhaps 10 instructors and 100 trainees, be instituted to determine how the instructors are likely to respond. A suggested set of instructions and TTTCL format is presented in Appendix II.

IV. Peer Ratings

A. General

Ever since Wherry et al (1949), and later Hollander (1956) showed that peer nominations predict the effectiveness of future military leadership rather than officer candidate popularity, this rating technique has been used, as pointed out earlier, as a successful approximation of ultimate performance criteria in a diversity of training and job situations. It has been further shown that peer ratings are not adversely affected by rater intelligence (Doll, 1963) or the stability of the situation in which the peer nominations are made (Borgatta et al, 1963).

With respect to reliability, one factor analytic study (Wolins, 1956) has reported that peer ratings made by 974 enlisted men in a gunnery training school had reliabilities in the high .80's. Although Wolins suggests that these reliabilities, because of the method used to approximate them, may be spuriously high other authors (Medland et al, 1964; Suci et al, 1954; and Mayo, 1954) have reported equally encouraging reliabilities under considerably less than favorable conditions.

B. The Use of Peer Nominations to Evaluate Trade and Technical Course Trainees

The use of the peer nomination technique is rather straightforward. All that is necessary is a roster of participant names and specific rater instructions. This point is important since it has been shown (deJung, 1964) that response set generated by different instructions influence rater nominations. While it is recommended that the peers, in this case classmates, be acquainted with each other for at least 6 weeks this is not an absolutely necessary condition. In this regard, it is interesting to note that Vielhaber et al (1965) have reported that impressionistic ratings made after 20-35 seconds were highly correlated with peer nominations made after several weeks of intensive observation!

It is further recommended that only favorable, i.e., positive, ratings be obtained since as found by Webb (1955) the use of negative ratings yielded results nearly identical with that found with positive ratings. More importantly, the use of only positive ratings precludes the possibility of rater resistance which is sometimes found when negative ratings are required.

Two other factors must be considered when using the peer nomination technique. They are the optimum group size and the number of peer choices that should be made and methods for deriving standard scores.

With respect to group size, it appears that 12-15 participants is the optimum number. If at all possible, the group should not be less than 10 or more than 20. As a rule of thumb, the participants should be required to nominate approximately 20 percent of the group members. The lower limit, regardless of group size, is 2 nominations. For example, in a group of 18 participants, the raters should make 4 nominations. In a group of 7 participants 2 nominations would be required. It should be noted that a participant should be allowed to nominate himself and that he not be required to identify himself. A sample roster and suggested instructions is prescribed in Appendix III.

A peer nomination participant's (trainee's) final score will be the number of ratings he receives. However, to compare his relative standing (intermediate criterion score) with other trainee's it becomes necessary to convert his final raw score to a standard score. This is usually accomplished by first assigning a rank to his raw score (other trainee's in the group are, of course, also assigned ranks) and then converting the assigned rank to a Hull score (see Ghiselli, 1957, p. 99) or some other conventional standard score. A more sophisticated technique which is particularly useful with groups of unequal size is described by Willingham (1959) and is recommended if the size of the peer nomination groups in the study vary significantly. Group size may be thought to vary significantly if more than 50 percent of the groups vary from each other by more than 3-5 participants.

APPENDIX I.

149 Items Appropriate for Inclusion in the Trade and Technical Trainee Check-List

2
8
11
15
24
32
36
49
50
59
66
84
88
93
99
112
116
146
160
161
172
176
193
208
210
225
246
264
277
278
326
327
341
367
370
451
472
498
500
546
595
596
646
665
666
691
712
713
746
758
763
774
776
798
818
819
851
852
853
854
872
882
892
908
909
910
919
928
944
949
962
963
964
985
1004
1023
1024
1032
1033
1035
1044
1045
1052
1057
1061
1062
1067
1091
1098
1103
1109
1120
1136
1137
1149
1153
1165
1183
1185
1211
1223
1236
1258
1272
1277
1286
1300
1301
1317
1336
1355
1356
1390
1394
1414
1415
1435
1459
1484
1499
1500
1535
1557
1575
1602
1632
1649
1669
1684
1702
1703
1716
1736
1758
1797
1832
1861
1862
1903
1921
1928
1936
1942
1960
1967
1968
1981
1987
1997

APPENDIX II.

Suggested Rater Instructions and Rating Format for the Trade and Technical Trainee Check-List

TRADE AND TECHNICAL STUDENT DESCRIPTION CHECK-LIST

General

Please fill in the following information then read the instructions below before completing this check-list. Please print all information.

Student's Name ______________________________ Date ____________
(Last) (First) (Middle Initial)

Institute or Community College ______________ Location ____________

Trade or Technical Program ______________________________

Instructor's Name ______________________________
(Last) (First) (Middle Initial)

Instructions

The following statements have been developed to describe trainees in trade and technical education courses. Some statements are more positive than others, but they are not presented in any particular order. Please read all the statements and then check, in the space provided, those statements which you feel best describe the student under consideration. You may check as many statements as you think necessary to accurately describe him.

_________ 1. Is highly efficient

_________ 2. Follows through beyond assignment

_________ 3. Is slow

.

.

.

_________ 50. Is quick to grasp information

APPENDIX III.

Sample Peer Nomination Roster and Ratee Instructions for Trade and Technical Trainees

Instructions

The list of names shown below includes all the students in your class. Assuming you were a supervisor and was given the responsibility to hire three (this number will vary depending on group size) people to work in your company or shop, which three men listed on the roster would you hire? Please indicate your choices by putting a check mark in the space provided. Do not choose more than three men. You may choose yourself if you wish. You do not have to sign this sheet or identify yourself in any way.

CLASS ROSTER

__________ Adams, C. H.

__________ Bird, C.

__________ Brown, H. Z.

__________ Clifford, J. J.

__________ Frank, A. A.

__________ Getz, S.

__________ Gillespie, D.

__________ Goodwin, B.

__________ Mulligan, J.

__________ Smith, H. W.

__________ Soames, M.

__________ Winding, K.

REFERENCES

Barrett, R. S. Performance Rating. Chicago: Science Research Associates, Inc., 1966.

Borgatta, E. F. & Sperling, B. C. "A Note on the Stability of Peer Judgments in Independent Situations," J. of Psych. Studies, 1963, 14(1): 45-47.

Brogden, H. E. & Taylor, E. K. "The Theory and Classification of Criterion Bias," Educ. and Psych. Meas., 1950, 10: 159-186.

Bell, J. R. Prediction of Success in Industrial Drafting. Unpublished Master's Dissertation, University of California at Los Angeles, 1949.

Broadhurst, J. C. A Differential Prediction of Success in Vocational-Technical and Vocational-Industrial Education. Unpublished Doctor's Dissertation, New York University, 1949.

deJung, J. E. "Effects of Rater Frames of Reference on Peer Ratings," J. Exper. Educ., 1964, 33(2): 121-131.

Doll, R. E. & Longo, A. A. "Improving the Effectiveness of Peer Ratings," Personnel Psych., 1962, 15(2): 215-220.

Fisk, D. W. & Cox, Jr., J. A. "The Consistence of Ratings by Peers," J. Appl. Psych., 1960, 44: 11-17.

Flanagan, J. C. "Critical Requirements: A New Approach to Employee Evaluation," Personnel Psych., 1949, 2: 419-425.

Fleming, J. W. "Predicting Trade School Success," Industrial Arts and Vocational Education. 1938, 27: 315-318, 365-367, 422-426.

Fuchs, E. R., Woods, I. A., & Harper, B. P. Prediction of Job Success in Eight Career Ladders. Adj. Gen. Office, Personnel Res. Branch, Report No. 997, February, 1953.

Ghiselli, E. E. & Brown, C. W. Personnel and Industrial Psychology. New York: McGraw-Hill Book Company, 1955.

Goertzel, V. An Objective Rating Form for Job Success. Unpublished Master's Dissertation, University of California at Berkeley, 1941.

Hayes, R. B. "A Way to Measure Teaching," J. of Educ. Research, 1963, 57(1): 47-50.

Hollander, E. P. "The Friendship Factor in Peer Nominations," Personnel Psych., 1956, 9: 435-447.

__________________. "Validity of Peer Nominations in Predicting a Distant Performance Criterion," J. Applied Psych., 1965, 49(6): 434-438.

Kirkpatrick, D. L. "Techniques for Evaluating a Training Program," J. Amer. Soc. for Training and Develop., 1959.

Klieger, W. A., deJung, J. E. & Dubuisson, A. U. "Peer Ratings as Predictors of Disciplinary Problems," USA Personnel Res. Office, OCRD, Tech. Res. Note 124, 1962.

Kanuft, E. B. "Construction and Use of Weighted Check-List Rating Scales for Two Industrial Situations," J. Appl. Psych., 1948, 32: 63-70.

Mayo, R. J. Relationships Between Length of Acquaintance and Nature of Trait Rates and Agreement Between Raters, USAF Personnel Training Res. Cent. Bull. No. 54-55, 1954.

Medland, F. F. & Olans, J. L. Peer Rating Stability in Changing Groups. USAPRO Tech. Res. Note No. 142, 1964.

Nagle, B. F. "Criterion Development," Personnel Psych., 1953, 6: 271-283.

Novak, B. J. & Scheuhing, M. A. "Predicting Success in High School Industrial Courses," Industrial Arts and Vocational Education, 1951, 40: 391-394.

Patterson, C. H. "Predicting Success in Trade and Vocational School Courses: Review of the Literature," Educ. and Psych. Meas., 1956, 16(3): 353-395.

Prien, E. P. & Lee, R. J. "Peer Ratings and Leadership Group Discussions for Evaluation of Classroom Performance," Psychol. Reports., 1965, 16(1): 59-64.

Richardson, M. W. & Kuder, G. F. "Making a Rating Scale that Measures," Personnel J., 1933, 12: 36-40.

Suci, G. J., Vallance, T. R. & Glickman, A. S. An Analysis of Peer Ratings: I. U.S. Bur. Nav. Personnel Tech. Bull., No. 54-9, 1954.

Thorndike, R. L. Personnel Selection. New York: John Wiley & Sons, 1949.

Thurstone, L. L. "Attitudes Can be Measured," Amer. J. of Sociology., 1928, 33: 529-554.

Toops, H. A. "The Criterion," Educ. and Psych. Meas., 1944, 4: 271-297.

Torgerson, W. S. Theory and Methods of Scaling. New York: John Wiley & Sons, 1958.

Uhrbrock, R. S. "Standardization of 724 Rating Scale Statements," Personnel Psych., 1950, 3: 285-316.

______________________. "2000 Scaled Items," Personnel Psych., 1961, 14(4): 375-420.

Vielhaber, D. P. & Gottheil, E. "First Impressions and Subsequent Ratings of Performance," Psychol. Reports, 1965, 17(3): 9-16.

Webb, W. B. "The Problem of Obtaining Negative Nominations in Peer Ratings," Personnel Psych., 1955, 8: 61-63.

Weitz, J. "Selecting Supervisors with Peer Nominations," Personnel Psych., 1958, 11: 25-35.

Wherry, R. J. "Criteria and Validity," Handbook of Applied Psychology, Vol. I. New York: Rinehart & Company, 1950, pp. 170-177.

Wherry, R. J. & Fryer, D. H. "Buddy Ratings: Popularity Contest or Leadership Criteria?," Personnel Psycho., 1949, 2: 147-159.

Willingham, W. W. "On Deriving Standard Scores for Peer Nominations with Subgroups of Unequal Size," Psychol. Reports., 1959, 5(3): 397-403.

Wolins, L. The Validity of Peer Ratings as a Function of Rater Characteristics and Methods of Scoring. Unpublished Doctor's Dissertation, The Ohio State University, 1956.